DOCUMENT RESUME

ED 480 046 CG 032 619

AUTHOR	Geisinger, Kirk F.
TITLE	Testing Students with Limited English Proficiency.
PUB DATE	2003-08-00
NOTE	13p.; In: Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators; see CG 032 608.
PUB TYPE	Information Analyses (070)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Educational Assessment; *Educational Testing; *English (Second Language); *Limited English Speaking; *Population Trends; Test Construction; Test Use; *Testing Problems

ABSTRACT

Considerable testing occurs in the schools and in related educational settings. Schools are microcosms of society, and changes that affect society are also likely to affect the schools in similar ways. The composition of American society has been changing dramatically in recent years, and this particular change is one that has influenced schools considerably; its effect on testing is dramatic. This chapter describes some of the ways that testing needs to be considered in light of the population shifts that are occurring, beginning with a description of the extent of these changes, then a consideration of three areas of test use from the perspective of dealing with individuals whose native language is not English. (Contains 14 references.) (GCP)

Reproductions supplied by EDRS are the best that can be made
from the original document

ED 480 046

Testing Students With Limited English Proficiency

By
Kirk F. Geisinger

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CG032619

BEST COPY AVAILABLE

Chapter 11
Testing of Students with Limited English Proficiency

Kurt F. Geisinger

Considerable testing occurs in the schools and in related educational settings. Schools are microcosms of society, and changes that affect society are also likely to affect the schools in similar ways. The composition of American society has been changing dramatically in recent years, and this particular change is one that has influenced schools considerably; its effect on testing is dramatic. This chapter describes some of the ways that testing needs to be considered in light of the population shifts that are occurring, beginning with a description of the extent of these changes, then a consideration of three areas of test use (as described in Geisinger, 2002) from the perspective of dealing with individuals whose native language is not English.

Population Shifts in American Society

Many (e.g., Eyde, 1992) have noted changes in American society. The predominant change is an increase in groups that do not speak English. As discussed in the following section, this change is due to both immigration and increasing birth rates.

Changes in the Population as a Whole

According to the U.S. Census Bureau, the population of the United States was 275 million in 1995 and is expected to grow to 300 million by 2010, and to 338 million by 2025. From July 1, 1995, until July 1, 2000, the United States population grew by 12 million people, approximately 12.5 percent. This growth comes from two primary sources: immigration and increasing birth rates. Both of these factors are leading to increases in the numbers of language minorities in the United States, and this group is growing at rates faster than the rest of the population. Approximately 2.8 million of the increase from 1995 to 2000 emerged from immigration and of these, approximately 43 percent were Hispanic; 25 percent were White, not Hispanic; 24.5

percent were from Asia; and some 7 percent were Black, not Hispanic. The majority of the White, not Hispanic group came from Eastern Europe and the majority of the Asian group came from Southeast Asia. Thus, virtually all these immigrants are coming from countries where English is not spoken, or is not a primary language. The majority of the increases over this five-year period, however, occurred due to differential birth rates, that is, rates that differ by ethnic group membership.

On July 1, 2000, the U.S. Census Bureau estimated the ethnic breakdown of the United States population (rounded to the nearest whole percentage) as follows (Geisinger, 2002):

70 percent White, not Hispanic
12 percent Hispanic American
13 percent African American
4 percent Asian American
1 percent Native American

The U. S. Census Bureau estimates the ethnic breakdown of the United States population by the year 2025 (again rounded to the nearest whole percentage) will be as follows (Geisinger, 2002):

62 percent White (a decline of 8 percent)
18 percent Hispanic-American (an increase of 6 percent)
14 percent African-American (an increase of 1 percent)
6 percent Asian-American (an increase of 2 percent)
1 percent Native American (no change)

Several types of population changes are occurring. Numbers of Hispanic Americans are increasing relative to the population as a whole, and it is estimated that by 2025, they will account for 66 percent more of the United States population, relative to their current status. Asian Americans too are growing rapidly in number and will increase by 50 percent. African Americans are growing by a more modest 8 percent. These gains are offset by a more than 11 percent decrease in the relative proportion of the largest group: Whites who are not Hispanics. Therefore, the three largest minority groups are all increasing, with Hispanic Americans and Asian Americans increasing most rapidly. Whether schools are ready or will be ready to accommodate this large and increasing number of language minorities is not yet clear.

Changes in the Schools

A large and increasing group in United States schools is composed of those students whose native language is not English. This group is

frequently known as limited English proficient students, or LEP students. Determinations must be made as to whether these individuals should be educated in English, their home language, or a combination of the two, as is often found in bilingual education. From a psychometric perspective, the testing of these individuals represents a thorny problem. If they are tested in English, they may not be able to show optimally what they know and can do. On the other hand, it is pragmatically difficult to build tests that can assess these students in their home languages—impossible in many school districts and states where more than 100 different languages may be spoken in homes.

LEP students currently comprise some 14 percent of the total test-taking population in our nation's schools, with approximately 75 percent of these students being Hispanic. Of the remaining 25 percent of LEP students, approximately 50 percent are Asian American, primarily Chinese, Vietnamese, and Korean.

Of the Hispanic students, more than 50 percent speak English at home, some 25 percent speak mostly Spanish at home, and 17 percent report speaking English and Spanish equally often at home (National Center for Education Statistics, 2002). The mother's place of birth is the strongest predictor of the Hispanic student's primary language. The language that is spoken in the home of Hispanic students is also closely related to their educational level. For example, "49 percent of the Hispanic students who spoke mostly Spanish at home had parents with a high school education, compared with 83 percent who spoke mostly English at home" (National Center for Education Statistics: Condition of Education, Indicator 6, pp. 1–2). Over the 27 years from 1972 to 1999, the percentage of Hispanic students in the schools has risen dramatically, paralleling the growth in the population as a whole, and there are large geographical differences reflected in the percentage of Hispanics enrolled in schools across the regions of our country. Throughout the entire country, the percentage of Hispanics in public education has risen from 6 percent in 1972 to 16.2 percent in 1999, an increase of 170 percent. At their most numerous, in the western part of the country, however, Hispanics made up 31 percent of the public school population in 1999, up from 15 percent in 1972. At the other extreme is the Midwest, where the percentage of Hispanic students was 6 percent in 1999, up from only 1.5 percent in 1972. Across the country, in 1993–94, 31 percent of Hispanic, Asian, or Native American children were classified as LEP students. Overall, the LEP population in American schools has experienced a 300 percent increase from the early 1990s into the beginning of the twenty-first century. Clearly, the schools are

facing the challenges of teaching students whose English is at best generally below that of the majority group, and at worst, very poor (U.S. Department of Education, 1997). These increasing numbers of LEP students demand changes to many aspects of the educational process, including testing.

Critical Psychometric Factors in Testing LEP Students: Culture and Language

The kind of increasingly diverse society that the American melting pot is places demands upon the professional testing community: companies, testing professionals (especially those who develop tests), and those who use the tests that are developed. A number of critical factors must always be considered in making all testing decisions. These include the regularly found differences among cultural and ethnic groups in test performance, especially on cognitive tests of ability and on measures of school achievement. Second, because tests, whether cognitive or of other types, are inherently behavioral samples, and because culture affects behavior, culture too affects test performance. In fact, if culture affects behavior relevant to the domain covered by a test, it must also affect test performance or the test itself would not be validly sampling the behaviors underlying the test. A third factor to consider is that most tests are language specific. Language is considered by many anthropologists to be one major factor inherent in culture, but only a single factor among others.

Determining the composition of the group to be tested is a preliminary consideration for anyone involved in the testing of groups of students or other individuals. Those who make decisions about testing must be aware of the number and size of varying cultural, language, and ethnic groups present in the targeted population. Such data may be acquired from local sources or from national groups, such as the U.S. Census Bureau. Researching the demographics of a group is time well spent.

Three Decisions in Testing

There are three decision areas related to testing that are greatly affected by the composition of the group to be tested. These three are the selection of the testing instrument, the administration of the test, and the use of the test. The last of these, test use, also subsumes test interpretation, as the proper use of test data first involves the appropriate

interpretation of test results. Each of these three testing concerns is addressed in turn below.

Test Selection

All individuals who decide what test to use are faced first with a simple question: whether to buy an existing measure or to build one. A variety of factors influence one or the other possibility. An argument for purchasing an existing measure is the fact that a test publisher, at least if the publisher is a major test publisher or a test publisher who specializes in the area covered by the test, generally can bring more research and other resources to the test development process. Included in the test development process is being up to date on the latest strategies of testing and current content. Similarly, such a publisher can also likely gather more extensive validation and normative data. Normative and validation information should be in the test manual, and potential test users are encouraged to contact the test publisher or even the test author if they need answers to specific questions. Normative and validation data are critical for proper test score interpretation and use. If the test has been available for a reasonable period of time, then potential test users can also read evaluations of the measure in sources such as the Buros *Mental Measurements Yearbook*; the Test Critiques series; and assessment-related journals, or in some cases, textbooks, such as Anastasi and Urbina (1997). Of particular interest to the thrust of this chapter is the necessity of considering not only the validity of the test, but also its validity when used with the language minority populations present in a particular setting. In the United States, a finding that validity data are consistent across groups means that a measure is valid for the Hispanic population as well as for the majority population. In specific settings, of course, other language groups may need to be considered.

When one chooses to develop one's own test, the standard factors involved in any test construction demand consideration. If the test is to be administered to and used with a linguistically diverse population, the questions one must ask about the test become much more complex. The ultimate questions that must be asked in any decision-making process relating to test selection and development are (a) is this measure valid for the use that is planned, and (b) is the test appropriate for all the groups involved? The former question requires the potential test user to decide whether there is evidence that the test can provide the kind of useful information that can enlighten decision-making in a particular context. (In the case of an admissions decision, for example, a valid test would provide information on which potential students are

most likely to succeed in the ensuing educational program and which are not. In the case of an achievement mastery test, a valid test would provide strong indications of the extent to which different students have in fact learned the material provided in the program.) While data supporting such a contention may emerge from a single study, it is more likely to come from a series of studies, which may or may not have been performed in sequence by the test developer or another test researcher. Such information is most commonly available either for the entire population or for the majority group within that population. Of considerable interest to those testing diverse populations, however, is how well the test works when used with subgroups of the population.

The second question is therefore somewhat more difficult. It relates to whether the kind of validation information called for is available for the varying subgroups within the population. Critical to the present discussion, of course, is whether this information is available for language minorities, in particular, the kinds of language minorities found in the setting of most interest to the potential test user. Simply put, the kinds of validation evidence that are employed to justify the use of a test with the entire population (or with the majority population) must also be present for all of the language minority groups.

Let us consider a few examples. Does a college admissions measure that predicts collegiate grades reasonably for students across the country also work when applied to Hispanics? Does it also work for recent immigrants whose English is quite weak? Does a measure of knowledge in history work for students across the country who have had college-preparatory courses in history throughout their high school education? That is, does it represent the information provided in the curriculum in a representative and fair manner? Does the same measure also fairly and accurately represent the curriculum of students who have been exposed to a bilingual curriculum, which includes some learning in English and some in their home language so that these students do not fall behind their peers as they "catch up" in English? Does it represent the courses taught in an inner-city school where multiple languages and cultures are present? For both of these types of tests, are they valid for individuals whose knowledge of English makes it difficult for them to read and comprehend the test questions as they are presented? Are they valid for individuals whose English mastery does not permit them to read the questions and the choices of answers and to respond to them as quickly as the majority group in our population? Test publishers who wish their tests to be used with linguistically diverse candidates should provide information supportive of positive responses

to the preceding questions. To be sure, however, such research is expensive, and only the largest of test publishers are frequently able to perform this research, regardless of its appropriateness and import.

A number of issues must be considered about an instrument that will potentially be used with language minority, or LEP, children. The issue of differential validity is paramount. The issue of whether the test is fair and unbiased is closely aligned with the validity issue. A third issue relates to norms; this topic is discussed in the treatment of test score interpretation and use. The final questions relate to whether there are forms of the measure more appropriate to LEP students (either in their home language or in an English-language reduced version) or whether there is interpretative information so that test users working with LEP children can effectively understand the meaning of these students' scores. (This last type of information is also closely related to the question of validity.)

The question of differential validity is most typically seen in the case of a test that is justified on the basis that it predicts a criterion. Differential validity is established if the test does not predict comparably for a minority group as it does for the majority group. Differential validity can extend to other forms of validity, however. If two groups (the majority group and a minority group) receive very different instruction in schools, for example, a test that covers only the content presented to the majority group could be seen as having differential content validity. Ultimately, the question of differential validity relates to whether the results of testing are equally meaningful for all groups. In the case of LEP students, such questions are critical, for international students have almost assuredly been exposed to different content in their instruction, and even those in the United States may have experienced somewhat different instruction, as for example, if they are in bilingual or remedial instruction.

One type of fairness is actually an assessment of differential validity. Such analyses normally consider the test as a whole. If a test is differentially valid but is used as if it is not, then at least one group will likely receive inappropriate results. It is also possible to consider discrete components of tests, especially individual test questions, to determine if they contribute to the biased nature of a test. Such analyses are called *differential item functioning analyses*, or *dif analyses*. (See Berk, 1982; Embretson & Reise, 2000; or Wasserman & Bracken, 2002, for in-depth treatments of this topic.) Essentially, what dif analyses do is consider whether individual test items are differentially more difficult relative to other items on the test for specific, identifiable subgroups in

the population. Such analyses are best performed during the test construction process so that items that do not function equivalently for all groups may be removed from draft versions of an examination. Those involved in the selection of a test are well advised to review the procedures used in the development of tests to see if dif procedures were employed and, in particular, if they were employed using the language minority subgroups to be tested.

Some tests are available in more than one language version, for example, in English and Spanish. Ideally, in such a case, the different forms have been developed and studied in ways to ensure their comparability. (See Geisinger, 1994, or Sireci, 1997, for considerations of some of the issues involved.) If, as is most commonly the case, a test is developed in one language and translated to a second, the term *adaptation* is used rather than *translation*. The reasoning behind this nomenclature is that changes in tests are not related only to language; culture too requires the original language form of a test to be changed to make sure that any references to aspects of culture are equivalent across the two forms. Such a process inevitably involves committee processes in which individuals who know about the content and constructs measured by the test, who are fluent in both languages, and who are knowledgeable about both cultures consider the test item by item to ensure that the two forms are indeed equivalent. A test that is available in more than one language obviously has advantages over one that is not. Nevertheless, the technical considerations that are involved in adapting a test from one language and culture to another are extensive and are infrequently performed in a superlative manner. A prospective test user must become familiar with the requirements involved in test translation and adaptation and inspect the procedures carefully before deciding to use a second language version of a test.

Test Administration

A few test administration issues are particularly relevant to the testing of LEP students. These include the use of second language forms, testing in English, and the sociocultural context of testing.

Before assessing Hispanic students with a test in either English or Spanish, one should make an assessment of each individual's relative language abilities. Although there may be circumstances in which one language needs be used instead of the other, there are also circumstances where the most valid assessment of what a student knows or can do is simply of more critical importance. In such a case, assessments of language competence are needed first. The level of language skill

typically required to respond to written test questions in English is quite high, and it is likely that many children whose home language differs from English, but who appear orally to be quite conversant in English (and even bilingual), cannot respond to the level of academic English required by a written examination. The timing of an examination may also be a concern, because their speed of functioning in their second language is likely to be much reduced. An assessment of relative language skills permits a determination of the language in which to test the LEP student using the proper language form of the examination.

If a language test indicates that a LEP child should be tested in English, or if no second language version of the test (or a comparable test) is available, then the child may need to be assessed in English. In such a case, it is possible that interpretations specific to those whose home language is not English may be needed. Such interpretations will likely be based on validation research using students with similar language skills and normative data using comparable groups. It is possible, for example, that a test score may have a different meaning for a student whose native language is English than for one whose native language is Spanish. This demarcation may be especially true if English has significant weight on the test, even if that is not what is intended to be measured by the test (as in the case of a master test of mathematics using many word problems, an essay test of American history, or a scale measuring test anxiety). In such an instance, the impact of language ability on the resultant scales is actually a source of test invalidity, because it reflects something other than what the test was intended to measure.

A test user should determine whether it is appropriate in a given context to use norms for an entire group (that is, the whole population tested) or for the specific group, of which the individual is a member (such as Hispanic children of a given age). Differing rationales argue for each in given contexts, and no general rules are advanced here for making this determination. One does need to determine the extent to which children with backgrounds and language skills similar to those being assessed were included in the reference or norm group. One should also determine whether norms for language minority children are also available, and if they are, whether the child or children being assessed are comparable to those in that specific norm group. Such information can greatly aid in the interpretation of a child's performance. In the same sense that a good test administrator should first assess an LEP child's language skills, the test administrator should also consider assessing the child's acculturation. (A brief discussion of acculturation

and its impact on test scores follows in the section on test interpretation and use.)

Anastasi and Urbina (1997) describe the transcultural context that sometimes occurs in testing situations. An example of a transcultural context is when a middle-aged White psychologist administers an individual test to a Hispanic youngster who has not had significant exposure to such individuals. Novelty, fear, and cultural factors can influence the child's performance; although such factors have generally not been found in investigations, they have occasionally been noted, and test administrators should be alert for such possibilities.

Test Interpretation and Use

Most professional test users determine the meaning of scores using validity and norms. Norms help us to interpret where an individual's performance places him or her relative to that person's particular reference group. Sandoval (1998) has called for what he terms "critical thinking in test interpretation." As such, Sandoval calls for those interpreting the test performance of students to examine carefully their preconceptions and the factors they use in explaining performance. Stereotypes are one such possible explanation of behavior against which testing professionals need to guard. Sandoval recommends using the factors that have been properly shown to aid in test score interpretations—such as test validity, norms, base rates, looking at extra-test behavior in addition to test scores and performance—and considering a longer time period than just the testing itself in making proper interpretations of test results.

Test users can follow general principles for permitting culture and cultural differences to influence interpretations of test performance (see Geisinger, 2002). It is particularly important that those using tests understand how members of specific groups tend to perform on given assessments in specific domains. The Office of Ethnic Minority Affairs of the American Psychological Association (1993) has provided guidelines for test interpretation. One is especially relevant. Guideline 2d states, "Psychologists consider the validity of a given instrument or procedure and interpret resulting data, keeping in mind the cultural and linguistic characteristics of the person being assessed. Psychologists are aware of the test's reference population and possible limitations of the instrument with other populations" (p. 46).

The acculturation of culturally diverse individuals being tested should be assessed, just as their language skills should be. CuÈllar (2000) portrays culture as mediating relationships between personality

and behavior. That is, one needs to consider the culture from which an individual comes as part of an interpretation and attribution of his or her behavior, including behavior on tests. Acculturation occurs as one learns about and changes in conformance to a new culture. One's learning English after coming to the United States, for example, is one type of acculturation. Test results should be considered in light of the degree to which an individual who has come to this country has become acculturated. (See Geisinger, 2002, for a brief overview of acculturation issues in testing and MarÌn, 1992, and CuÈllar, 2000, for good summaries of issues involved in the assessment of acculturation.)

Conclusion

Our society is changing rapidly. These changes include dramatic changes in the numbers of LEP children in the schools. This influx affects testing. If the acculturation and English proficiency of linguistic minorities are high, tests are likely to be used effectively. To the extent that these factors are not high, however, difficulties often arise. This chapter has presented some information that should help test users in deciding whether to build or select a test to be used with this population, to decide which test to select, to administer the test properly, and to interpret scores accurately. Because these issues are so complex, only high points of the issues involved were mentioned. Test users who work with linguistically diverse populations need to be most concerned with validity, and they need to study test manuals and validation reports carefully to determine whether the tests are appropriate for the populations with which they work. They also need to consider normative information and research on the use of the instruments with the appropriate populations. Caution is, however, the overarching order of the day.

References

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Berk, R. A. (Ed.) (1982). *Handbook of methods for detecting test bias.* Baltimore, MD: Johns Hopkins University Press.

Cuellar, I. (2000). Acculturation as a moderator of personality and psychological assessment. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment* (pp. 113–130). Mahwah, NJ: Erlbaum.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Eyde, L. D. (1992). Introduction to the testing of Hispanics in industry and research. In K. F. Geisinger (Ed.), *Psychological testing of Hispanics* (pp. 167–172). Washington, DC: APA Books.

Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6,* 304–312.

Geisinger, K. F. (2002). Testing the members of an increasingly diverse society. In J. F. Carlson & B. B. Waterman (Eds.), *Social and personality assessment of school-aged children: Developing interventions for educational and clinical use* (pp. 346–364). Boston: Allyn and Bacon.

Marin, G. (1992). Issues in the measurement of acculturation among Hispanics. In K. F. Geisinger (Ed.), *Psychological testing of Hispanics* (pp. 235–252). Washington, DC: APA Books.

National Center for Education Statistics, Office of Educational Research and Improvement. (2000). Condition of Education: Indicator 6. Retrieved January 18, 2003, from nces.ed.gov/programs/coe/2000/section1/indicator06.asp

Office of Ethnic Minority Affairs of the American Psychological Association. (1993). Guidelines for providers of psychological services to ethnic, linguistic and culturally diverse populations. *American Psychologist, 48,* 45–48.

Sandoval, J. (1998). Critical thinking in test interpretation. In J. Sandoval, C. L. Frisby, K. F. Geisinger, J. C. Scheuneman, & J. Ramos Grenier (Eds.), *Test interpretation and diversity: Achieving equity in assessment* (pp. 387–402). Washington, DC: APA Books.

Sireci, S. G. (1997). Problems and issues in linking tests across languages. *Educational Measurement: Issues and Practice, 16* (1), 12–19.

U.S. Department of Education, National Center for Education Statistics. (1997). *A profile of policies and practices for limited English proficient students: Screening methods, program support, and teaching training.* (SASS 1993–94). NCES 97–472, by M. Han, D. Bakers, & C. Rodriguez. Washington, DC: Author.

Wasserman, J. D., & Bracken, B. A. (2002). Selecting appropriate tests: Psychometric and pragmatic considerations. In J. F. Carlson & B. B. Waterman (Eds.), *Social and personality assessment of school-aged children: Developing interventions for educational and clinical use* (pp. 18–43). Boston: Allyn and Bacon.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
X	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").